Sephora: Product Insights, Analysis & Sentimental Analysis

Project Overview:

The "Beauty Product Insights and Recommendations" project aims to analyze a large dataset of beauty products from the Sephora online store. By leveraging data analytics, machine learning, and data visualization techniques, this project aims to provide valuable insights into product popularity, customer sentiments, and potential areas for improvement. The project will culminate in the development of a recommendation system to suggest products to customers based on their preferences and body features.

Project Objectives:

Absolutely! Here are the project objectives presented in future tense:

- The project will analyse product categories, attributes, and variations to uncover emerging trends and patterns.
- Sentiment analysis will be conducted on user reviews to extract valuable insights into customer emotions and opinions.
- Interactive dashboards will be developed to visualize product insights, enabling data-driven decision-making.
- A recommendation system will be created to suggest personalized products based on individual customer preferences.
- Actionable recommendations will be provided to enhance Sephora's customer experience and product offerings.
-

Value of this project

The "Beauty Product Insights and Recommendations" project can create a significant impact and add substantial value to both customers and Sephora:

- ✓ Enhance Customer Experience: By analysing customer sentiments and preferences, the project can help Sephora gain a deeper understanding of their customers' needs and desires. This insight can lead to improved product offerings, more accurate recommendations, and a personalized shopping experience.

- ✓ Improve Product Strategy: The analysis of trends and patterns within the beauty product data can guide Sephora in making informed decisions about which products to promote, enhance, or introduce. This can optimize their product portfolio, leading to increased customer satisfaction and loyalty.

- ✓ Tailor Recommendations: The recommendation system developed as part of the project can enable Sephora to provide personalized product suggestions to each customer. This

not only enhances the shopping experience but also increases the likelihood of customers discovering and purchasing products that align with their preferences.

- ✓ Effective Marketing and Promotions: Insights gained from the project can help Sephora design targeted marketing campaigns and promotions. By understanding customer preferences and sentiment, Sephora can create more relevant and compelling messaging, ultimately driving higher engagement and conversion rates.

- ✓ Competitive Edge: In the highly competitive beauty industry, having data-driven insights and an effective recommendation system can set Sephora apart from competitors. This project can position Sephora as a brand that truly understands its customers and provides a superior shopping experience.

- ✓ Operational Efficiency: The project's findings can guide inventory management, allowing Sephora to stock products that are more likely to sell and reduce the risk of overstocking or understocking. This efficiency can lead to cost savings and improved resource allocation.

- ✓ Informed Decision-Making: The interactive dashboards and visualizations created during the project can empower Sephora's decision-makers with clear, actionable insights. This can lead to more strategic and well-informed business decisions across various departments.

- ✓ Continuous Improvement: The project's recommendations and insights can serve as a foundation for ongoing improvements. Sephora can regularly update their product offerings, marketing strategies, and customer engagement based on evolving trends and customer feedback.

About Dataset

This dataset was collected via Python scraper in March 2023 and contains:

Information about all beauty products (over 8,000) from the Sephora online store, including product and brand names, prices, ingredients, ratings, and all features.

User reviews (over 1 million on over 2,000 products) of all products from the Skincare category, including user appearances, and review ratings by other users This dataset contains multiple csv files such as:

- Product_info.csv that contains information about all products from the Sephora o nline store
- Five csv files that contains all reviews of the skincare products.

What inside this project?

1. **Data Collection and Preparation:**

   Downloaded the beauty product dataset from *Kaggle*. Cleaned and pre-processed the data by handling missing values, duplicates, and outliers. Ensure that the data is ready for analysis and modelling.

2. **Exploratory Data Analysis (EDA):**

   Performed exploratory analysis to understand the distribution of product categories, attributes, and variations. Visualize trends and patterns within the data using various plots, histograms, and charts. Identify the most popular product categories, best-selling products, and variations with high demand.

3. **Sentimental Analysis:**

   Extract user reviews and ratings from the dataset. Employ Natural Language Processing (NLP) techniques, including tokenization, stemming, and lemmatization, to pre-process text data. Utilize Tensor flow to train and test sentiment analysis models. Classify reviews into positive and negative sentiments, enhancing understanding of customer opinions.

4. **Actionable Recommendations:**

   Based on the insights gained from the analysis and the recommendation system, it will allow to recommend strategies to enhance customer experience, such as improving product descriptions, optimizing pricing, or introducing new product categories.

Results and Discussions

1. Brand Analysis:

- Visualized the distribution of brands based on 'brand name' using a bar plot.
- Calculated and compared the average 'loves count', 'rating', and 'reviews' for different brands.
- Identified the top brands with the highest average 'rating' or 'loves count'.

2. Performed Price and Variation Analysis

- Also I have shown the distribution of ‘price (USD)' using a histogram or box plot.
- Analysed how ‘price (USD)' varies across different 'variation type' or 'variation value' categories.
- Explored the relationship between ‘price (USD)’ and 'loves count' using a scatter plot.

3. Category Insights:

- Created a bar plot to show the distribution of products across 'primary category', 'secondary category', and 'tertiary category'.
- Compared the average 'rating' or 'loves count' for different categories.

4. Product Attributes:

- Visualized the distribution of 'size' using a histogram or box plot.
- Analysed the most common 'ingredients' or 'highlights' using word clouds or bar plots.
- Explored the relationship between 'size of products' and ‘price (USD)’.

5. Availability and Exclusivity:

- Compared the percentage of products that are 'limited-edition', 'new', 'online only', or 'Sephora exclusive'.
- Analysed how availability status ('out of stock') affects 'loves count' or 'rating'.

6. Child Products Analysis:

- Investigated the distribution of 'child count' (number of variations) using a histogram.
- Compare the average 'price (USD)', 'loves count', or 'rating' for products with different 'child count'.

7. Highlight Analysis:

- Analysed the most common 'highlights' or 'features' using word clouds or bar plots.
- Determine how the presence of specific 'highlights' relates to 'rating' or 'loves_count'.

8. User Interaction and Product Engagement:

- Explored how 'loves count', 'reviews', and 'rating' are related to each other.
- Investigated whether products with higher 'loves count' tend to have higher 'rating' or more 'reviews'.

9. Product Naming Analysis:

- Analysed naming patterns or trends in 'product name', such as common keywords or phrases.
- Investigated whether the length of 'product name' relates to 'loves count' or 'rating'.

10. Performeed Sentimental analysis

- Leveraging NLP, Tensor Flow, and Keras, an accuracy rate of 89% was attained in sentiment analysis on Sephora's reviews dataset.
- This will empowers data-driven decisions, helps to fosters customer engagement and fuels business growth.

11. Also Performed EDA and Insights Extraction on Merged Product and Reviews Dataset

- Conducted extensive Exploratory Data Analysis (EDA) on the combined product and reviews dataset, uncovering hidden patterns and critical trends.
- Utilized customer skin quality to tailor brand recommendations, enhancing personalized experiences.
- Identified top-rated and positively-reviewed products, harnessing sentiment analysis to showcase crowd favourites.
- Highlighted both popular and negatively-reviewed items, offering insights into potential areas of improvement and customer sentiment.